Semi-automated Ki67 Index Label Estimation For HE Images Classification

Dominika Petríková, Ivan Cimrák
University of Žilina
Žilina, Slovakia
dominika.petrikova, ivan.cimrak@fri.uniza.sk

***Abstract*—The quantification of Ki67 proliferation index conducted through immunohistochemistry staining holds significant importance in histopathology. However, this process has several limitations, including variability and subjectivity in interpretation, along with challenges in quantification, rendering the process time-consuming and costly. Consequently, the utilization of neural network models presents a promising avenue for enhancing this domain. Yet, the creation of a sizable, high-quality annotated dataset remains a laborious task for experts. In this paper, we propose a validation and subsequent improvements of previously suggested semi-automated approach for generating Ki67 scores from pairs of hematoxylin and eosin (HE) and immunohistochemical slides, aiming to reduce the reliance on expert intervention. This approach integrates image analysis techniques such as clustering and optimization for tissue registration. Through the proposed approach and modification of worklfow, aiming to reduce the variability of the quantification error on different whole slide images, we annotated HE patches and conducted multiple experiments to fine-tune ResNet models for predicting Ki67 scores from HE images.**

I. Introduction

Digital pathology and image analysis are crucial in diagnosing various diseases including cancer. This field relies on high-quality digital scans of tissue samples, typically obtained as whole slide images (WSIs) using advanced digital scanners. The evolution of these scanners has facilitated the generation of vast amounts of histopathology data, which can be leveraged by machine learning algorithms for numerous tasks, including tissue specimen classification [1], [2].

Histopathological analysis, particularly of malignant tumors, is typically conducted on thin sections (3-4µm) of formalin-fixed paraffin-embedded tissues, initially stained with hematoxylin and eosin (HE). This staining enables the basic evaluation of tumor morphology, encompassing parameters such as mitotic activity, invasion of adjacent structures, and tumor grading.

Certain tumor categories, such as neuroendocrine neoplasias of the gastroenteropancreatobiliary and respiratory tract, necessitate immunohistochemical (IHC) analysis of tumor proliferation activity as part of their grading. This analysis involves utilizing IHC antibodies against the proliferation factor Ki67, a nuclear protein associated with ribosomal RNA transcription expressed during the interphase of proliferating cells [3].

In clinical practice, consecutive tissue sections undergo both IHC and HE staining. This enables pathologists to assess various tissue characteristics within the same area across adjacent slides. Acquiring IHC staining is a conventional practice in clinical settings to ascertain tissue molecular details, yet it presents several drawbacks. IHC procedures are time-consuming, costly, and reliant on tissue handling protocols, as the outcomes are typically expressed through stain intensity, presence or absence of stain, or the proportion of cells exhibiting detectable stain intensity [4]. Numerous recent studies have demonstrated a correlation between HE and IHC stained slides of the same area [4]–[6]. Consequently, it should be feasible to predict the expression of specific proteins directly from HE slides.

Training a deep neural network requires an extensive collection of annotated images of high quality to serve as a training dataset. WSIs are often too large to be processed fully by neural networks, they must be segmented into smaller images, referred to as patches. The complexity of evaluating patches is significantly high both in terms of the effort required for creation as well as the expertise of the evaluator.

A. Related works

The topic of automatic or semi-automatic quantification of Ki67 proliferative expression has been addressed in several studies. The task of predicting Ki67 cell positivity from HE images was addressed in [7]. The authors utilized a fine-tuning approach on ResNet18 using cell-level annotations from HE images. They employed a point label approach to annotate cell patches based on homogeneous regions of Ki67 positivity or negativity. However, their method cannot be applied to heterogeneous images with mixed positive and negative cells.

Several studies [8], [9] have attempted to train a model that would automatically be able to detect and classify individual cell nuclei and determine a Ki67 index based on these counts. The drawback of this approach is the need for a well-annotated dataset to train the models. Authors in [10], [11] have addressed the problem of counting Ki67 positive and negative cells based on image analysis and unsupervised methods alone. They utilized color separation and thresholding to detect and classify nuclei. In addition, researchers in [12] have used hybrid cluttering based on segmentation. All of these approaches use the number of nuclei detected for quantification, which can be difficult to determine in cases where the tissue is very dense or the section is thick and it is very difficult to distinguish individual nuclei in clusters.

B. Contents of this work

The goal of our research is to create a classification model for the prediction of Ki67 index from an HE stained image from semi-automated annotated data. In our research, we apply image analysis methods to estimate Ki67 index for HE patches from pairs of adjacent HE and Ki67-stained tissue sections based on the area occupied by cells not cells count itself. Our proposed approach offers a significant advantage in that it necessitates minimal manual annotation by experts. The paper demonstrates our preliminary results.

The objective of this study is to validate our proposed semi-automated workflow for annotating HE patches by Ki67 expression and develop a neural network model capable of classifying Ki67 protein expression from HE images. Initially, we outline our proposed methodology for semi-automated dataset creation, followed by a presentation of experiments conducted in training multiple neural network models for classification.

The paper is organised as follows. Section II introduces our annotation process of dataset and model used for classification. Section III reports applied validation of dataset creation workflow together with our preliminary results in classification, and section IV concludes with future work.

II. Methods

A. Dataset construction

The data comprised of a 84 pairs of HE and Ki67-stained whole slide images of seminoma, a testicular tumor, sourced from the Department of Pathology, Jessenius Medical Faculty of Comenius University and University Hospital, Slovakia. To ensure tissue similarity, HE and Ki67 staining were applied to adjacent tissue sections. Despite slight disparities at the cellular level, we proceeded under the assumption that tissues from corresponding regions in both images shared similar characteristics.

Given the absence of labels in the dataset, it was imperative to annotate the data first. For this purpose, we adopted an enhanced semi-automated methodology inspired by in-press [13], involving three primary steps: tissue registration, clustering into primary colors, and quantification of the Ki67 index. Our scans contain images in 8 resolution levels, marked from 0 to 7 starting with the highest resolution. Due to computational constraints and the immense size of the scans, we processed images from level 1, containing the second-highest resolution. To align the images so that tissues from the same area are in the same position, predefined pairs of HE and Ki67 keypoints were used in conjuction with optimization method for finding best transformation parameters between them. After tissue registration, K-means clustering with respect to HSV (hue, saturation, value) colors of the pixels is applied to Ki67 stained image with resulting centroids representing ”typical colors in the slide”. Resulting centroids are subsequently divided into three categories: positive cells (brown colors), negative cells (blue colors) and background (white colors). In addition to clustering on HSV space, we applied K-means clustering to another two color spaces: RGB (red, green, blue) and Lab. However, according to elbow method, we obtain better results with HSV color space.

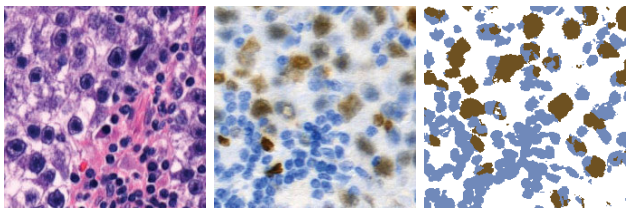
Ki67 indexes for HE patches were derived from corresponding Ki67 patches utilizing image analysis techniques. Subsequently, each computed Ki67 index was assigned as the label to the corresponding HE patch, taken from the same location. An illustrative example depicting an HE patch alongside its corresponding original Ki67 patch is presented in Figure 1. The third clustered Ki67 patch was generated as an outcome of the clustering. These clustered Ki67 patches were instrumental in estimating the Ki67 indexes, calculated as a ratio of brown to blue pixels.

B. ResNet18 model

Deep Learning has an important role in Machine Learning research by incorporating very deep neural networks for solving many issues, namely in image recognition domain. The basis of many deep learning architectures is the Convolutional Neural Network (CNN) [14], a model inspired by biological neural networks featuring localized connections and weight sharing.

Architectures of CNNs can be derived from established networks like VGG [15], Inception [16], or ResNet [17], or custom-designed to suit specific requirements. Each approach has its merits and drawbacks. Leveraging pre-trained networks offers the advantage of utilizing weights acquired from training on large-scale datasets like ImageNet [18], facilitating transfer learning and expediting training. Conversely, custom-designed networks offer flexibility, enabling tailored architectures suited to dataset characteristics, albeit with potential lower accuracy when training data are limited.

ResNet is a prominent deep learning architecture built of residual blocks displayed in Fig. 2. In general, ResNet is distinguished by its incorporation of ”shortcut connections”, which streamline optimization and enhance accuracy by accommodating increased network depth. Skip connections also called identity shortcut connections solve the problem of vanishing gradient and learning of identity function. In our study, we adopted a modified ResNet18 architecture for our CNN classifier. We replaced the classification segment of the original architecture by incorporating two fully-connected layers featuring 512 and 2 or 3 neurons (based on classification task), respectively. Additionally, a dropout with a value of 0.2 was applied to the fully-connected layer with 512 neurons.



(a) Original HE (b) Original Ki67 (c) Recolored Ki67

Fig. 1. Example of patches from dataset used to create annotations

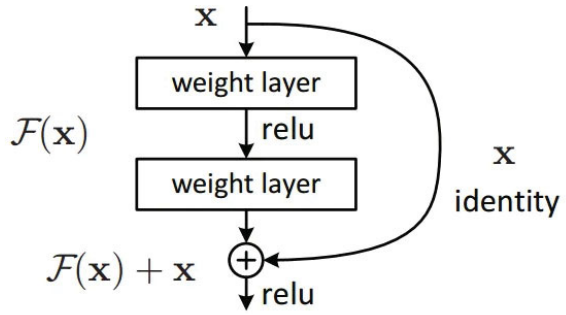


Fig. 2. Residual block architecture [17]

III. Results

Utilizing the methodology outlined in the Methods section, we generated an extensive dataset comprising over 90,000 annotated patches, each sized 224x224, extracted from specified regions of interest (ROIs) across 75 whole slide images. These ROIs were meticulously selected based on tissue quality, ensuring the exclusion of patches with black artifacts or blurriness from the dataset. To closely simulate real-world model deployment scenarios, we partitioned the dataset into a training set and two distinct validation sets (VSs). The first validation set was formed by segregating WSIs themselves, with seven WSIs earmarked for exclusive use in validation, mirroring the practical scenario where the model must predict values for new patients. Meanwhile, the second validation set comprised 10% of the patches extracted from the training set, facilitating ongoing training progress tracking and validation on data unseen but originating from tissues familiar to the model.

To facilitate machine learning-based analysis on HE and Ki67-stained slides containing testicular seminoma samples, in accordance with pathologist recommendations, we established three distinct thresholds for Ki67 expression: below 20%, between 20-50%, and above 50% Ki67 expression levels. A breakdown of patch counts in each class reveals a notable class imbalance: the below 20% class contained 90,000 patches, the 20-50% class comprised 8,500 patches, while the above 50% class encompassed fewer than 800 patches.

Given this substantial class imbalance, our initial approach involved binary classification with the first two classes, which had sufficient patch counts. To address the imbalance, we undertook dataset balancing through undersampling, randomly selecting approximately 10,000 patches from the below 20% and 20-50% categories for inclusion in the dataset.

Across all experiments, we employed the ResNet architecture pre-trained on ImageNet, specifically leveraging ResNet18. For model training, we utilized the SGD optimizer (Stochastic Gradient Descent) with both momentum 0.9 and weight decay 0.0001. All models underwent training for 100 epochs with a batch size of 64 and a learning rate set to 0.001, unless stated otherwise.

In conjunction with addressing data imbalance, we incorporated horizontal and vertical flip data augmentation during the training process. Several studies [19]–[22] have reported that when classifying HE images, it is advisable to do stain normalization of the data, as these can have a very heterogeneous range of colors, which makes it difficult for the models to be able to classify them. This is particularly applicable to cases where the data were collected from multiple laboratories or were acquired using different staining protocols, which is not our case. Despite this, when we look at our data, the color is very heterogeneous. Therefore, we tried to add color augmentation to the data with changing brightness, contrast, saturation and hue values. In this case, we let the model train longer and applied an early stopping method. We report the results obtained in Table I, where we can see that on one hand, any manipulation of the data improved significantly the accuracy of the models on the second VS. On the other hand, on the first VS color augmentation brought only small improvement and classification performance is still poor on new slides.

A. Quantification Validation

The estimation of quantification was validated on a small sample of data for which we have empirically evaluated ratios of positive cells directly from pathologists. This validation is discussed in more detail in [23]. The weakness of this validation is that the data to which we apply the semi-automated approach are significantly different from the data on which the procedure has been validated as can be seen in Fig. 3. This is due different tissue handling protocol used during staining. For this reason, it was also necessary to perform the validation directly on the data that we want to further use for classification purpose.

The SlideViewer medical software, utilized by pathologists in clinical practice, offers numerous additional functionalities for tissue analysis including quantification of tissue IHC positivity displayed on the screen. Leveraging this, we acquired a substantial amount of relatively well-annotated data suitable for validating our proposed Ki67 quantification method. Following parameter fine-tuning by experts to ensure accurate ratio quantification, this method can be applied to the actual tissue displayed on the screen, with results exportable. In this way, we obtained 16 annotated cutous from 6 scans.

After applying the clustering and quantification process to each cutout separately (as before in dataset construction), we obtained an average estimation error of 22% with minimum error 1% and maximum error 50%. The high variance between the errors on individual images confirmed our assumption that the low accuracy of the models on the first VS is due to the clustering and centroid extraction being performed on each slide separately, so the error is slide-specific.

In order to reduce the dependence of the ratio approximation error on a particular slide and lower slide-specific bias, we decided to use the annotations obtained from the experts to fine-tune clustering, by obtaining centroids from merged images to be used to recolor and estimate all slides. We divided

TABLE I. RESULTS OF RESNET18 MODEL WITH DIFFERENT COLOR AUGMENTATION APPLIED FOR BINARY CLASSIFICATION ON DATA WITH SLIDE-SPECIFIC CLUSTERING

Category split	First validation set			Second validation set		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Original	0.5497	0.5074	0.5097	0.7973	0.7973	0.7979
Brightness, saturation, hue, contrast	0.6180	0.5379	0.5454	0.8530	0.8547	0.8502
Brightness, saturation	0.6271	0.5325	0.5370	0.8348	0.8348	0.8348

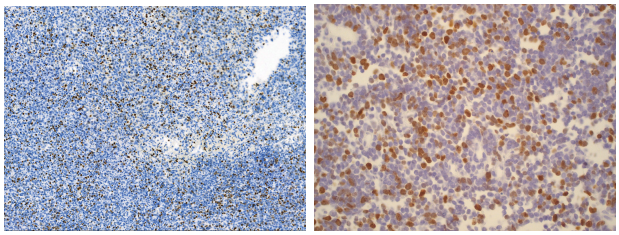


Fig. 3. Comparison of Ki67 stained tissue image used for classification (left) and Ki67 stained image used for validation (right)

the slides from which we had annotated cutouts into ”training” and ”validation”, similar to training neural networks. From the training annotated cutouts we selected 10, which we combined into a single super image over which clustering with K-means method was performed with K set to 12, 18, 24 and 30. We then divided the obtained centroids into 3 categories, experimenting with multiple combinations of assigning centroids, which were difficult to assign into one category at first sight. In this way, we were able to obtain a distribution of centroids that produced the lowest error of 13% on both validation and training cutouts when estimating the Ki67 positivity ratio. We used these centroids to recolor the entire dataset and obtain new annotations for the patches that were used to train the new models described in the classification subsection.

B. Registration Validation

Our proposed data annotation method comprises two primary phases: registration and quantification. While we were able to validate the quantification of the Ki67 ratio through manual annotations and estimations using the medical software described in previous section, assessing the accuracy of registration remained challenging, with visual inspection being the sole validation method available. Consequently, we sought to reinforce the correctness of the registration process by leveraging the accuracy of the classification model.

Given that each pair of patches corresponds to tissue with similar properties, we expected the CNN to learn the relationship between them not only on the training set but also on the validation set. To test this hypothesis, we compared the accuracy of two models. The first model was trained on an existing dataset by default, while the second model was trained on a randomly generated dataset.

For the random dataset, we adopted the distribution of Ki67 expression ratios from each scan and randomly assigned annotations to the patches to match the existing distribution. In essence, we randomly paired HE and IHC patches from each scan. A model trained on such a dataset should exhibit lower validation accuracy, as the annotations for the validation data are randomly assigned, and the neural network cannot effectively model this randomness.

It’s worth noting that our dataset is highly unbalanced, with patches in the category below 20% outnumbering other categories. The majority of patches fall within the category below 5%. Therefore, we initially balanced the dataset before performing random assignment. This approach ensured that despite the random assignment, a substantial portion of the data would still be correctly categorized, potentially allowing the model to detect this partial information.

The outcomes of both models are outlined in Table II. It is evident that the first model attained a notably higher accuracy, whereas the second model’s accuracy is comparable to random guessing. Consequently, the accuracy on both validation sets is considerably lower compared to a model with annotations estimated from registration. This serves as compelling evidence for us to affirm the accuracy of the procedure.

C. Classification on shared centroids annotated patches

After applying common centroids to recolor and annotate the entire dataset, both the number and distribution of patches in each category changed. From the original values, the numbers of patches changed to the following: in below 20% class there remained 33 500 patches, between 20%-50% was 17 700, and in the last category above 50%, was 4 200 patches. Since patches count in the last category has increased, we used combination of undersampling and oversampling method for dataset balancing, with undersampling two largest classes and oversampling the minority class. With the increased numbers in the class above 50%, we were able to test multiclass classification over the full range of the ratio, not only binary classification. Results of both classification: binary (below 20% and between 20-50%) and multiclass are shown in Table III and Table IV respectively. On the binary classification results, we can observe a decrease on the second VS and, conversely, an increase on the first VS, which is the phenomenon we predicted. Again, higher accuracy was achieved by the models where color data augmentation was used, but the increase in accuracy on the second VS is not high enough for the model to be used in practice. Moreover, on the multiclass classification we can observe comparable accuracy to the binary classification on the second VS, however the accuracy on the first VS is significantly lower indicating the higher complexity of such a task on new data.

Since the estimation error of the accuracy 13% is quite high despite efforts to reduce it, this may confound the model for data that are at the edges of the category intervals. Therefore,

TABLE II. ACCURACY OF RESNET18 MODEL TRAINED ON DATASET WITH CORRECTLY AND RANDOMLY ASSIGNED LABELS

Dataset	First validation set			Second validation set		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
correct	0,6122	0,5149	0,5169	0,8864	0,8864	0,8864
random	0,4929	0,3184	0,3117	0,5205	0,5196	0,5196

TABLE III. RESULTS OF RESNET18 MODEL WITH DIFFERENT COLOR AUGMENTATION APPLIED FOR BINARY CLASSIFICATION ON DATA WITH SHARED CLUSTERING

Category split	First validation set			Second validation set		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Original	0.5606	0.5688	0.5699	0.7745	0.7749	0.7731
Brightness, saturation, hue, contrast	0.5949	0.5912	0.5933	0.7702	0.7707	0.7704
Brightness, saturation	0.5731	0.5885	0.5883	0.7863	0.7883	0.7852

TABLE IV. RESULTS OF RESNET18 MODEL WITH DIFFERENT COLOR AUGMENTATION APPLIED FOR MULTICLASS CLASSIFICATION ON DATA WITH SHARED CLUSTERING

Category split	First validation set			Second validation set		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Original	0.4628	0.3989	0.3965	0.7756	0.7700	0.7747
Brightness, saturation, hue, contrast	0.4218	0.3611	0.3860	0.7869	0.7830	0.7859
Brightness, saturation	0.4280	0.3748	0.3692	0.7891	0.7847	0.7885

in the next experiment, we adjusted the data and omitted 20% between each interval. Hence, the classification classes changed to below 10% between 30-50% and above 70%. Our assumption is that this should increase the accuracy of the models, as it should eliminate situations where patches with similar ratios will be assigned in two different categories due to estimation error.

However, this assumption as we can see in the Table V and the Table VI proved to be true only on the second VS, where the accuracy increased significantly from 0.77 to about 0.9. On the other hand, the performance of the models on the first VS did not change, which suggests that the error of the annotation method still has a high variance and in the case of slides where the error is larger than half of the omitted interval, it can still confuse the model. Therefore, these preliminary results proves that this problem needs to be addressed more deeply in the future.

IV. CONCLUSION

In this study, we validated our semi-automated workflow for Ki67 quantification consisting of two main steps: tissue registration and quantification. We utilized ResNet18 model to classify HE patches into Ki67 index categories thus obtained resulting in model with good performance on validation patches, extracted from training slides not present in training set. However, due to low accuracy on new slides, we proposed an enhancement to the previous annotation extraction technique, diverging from "in press" [13] by extracting shared centroids used for all slides in order to reduce per slide estimation error variability. Although this has slightly increased the classification accuracy of the binary model in particular, the performance is still not sufficient and annotation error is still high. Despite efforts to increase model accuracy by omitting boundary intervals between categories, this was only successful on one validation dataset. Therefore, the annotation workflow quality still needs to be enhanced.

In our future research, we will improve the quality of the Ki67 index estimation by closer examinating color properties of the IHC images taking into account the positivity intensity of the cell. Thus try to reduce the variability of the estimation error, which should yield an increase in classification ability on the new data. In addition to modifying the dataset creation, we will employ neural network explainability methods to further investigate which parts the model gives more importance to.

ACKNOWLEDGEMENTS

This research was supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic under the contract No. VEGA 1/0369/22.
Gratitude is extended to K. Tobiášová and L. Plank from Department of Pathology, Jessenius Medical Faculty of Comenius University and University Hospital, Martin, Slovak Republic for their collaboration in the preparation of the dataset.

REFERENCES

[1] P. W. Hamilton, P. Bankhead, Y. Wang, R. Hutchinson, D. Kieran, D. G. McArt, J. James, and M. Salto-Tellez, "Digital pathology and image analysis in tissue biomarker research," Methods, vol. 70, no. 1, pp. 59–73, 2014, advancing the boundaries of molecular cellular pathology. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1046202314002370

[2] L. Pantanowitz, "Digital images and the future of digital pathology: From the 1st digital pathology summit, new frontiers in digital pathology, university of nebraska medical center, omaha, nebraska 14-15 may 2010," Journal of Pathology Informatics, vol. 1, no. 1, p. 15, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2153353922001079

[3] J. Bullwinkel, B. Baron-Lühr, A. Lüdemann, C. Wohlenberg, J. Gerdes, and T. Scholzen, "Ki-67 protein is associated with ribosomal RNA transcription in quiescent and proliferating cells," J. Cell. Physiol., vol. 206, no. 3, pp. 624–635, Mar. 2006.

TABLE V. RESULTS OF RESNET18 MODEL WITH DIFFERENT COLOR AUGMENTATION APPLIED FOR BINARY CLASSIFICATION ON DATA WITH SHARED CLUSTERING WITH OMITTED INTERVALS BETWEEN CATEGORIES

Category split	First validation set			Second validation set		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Original	0.5860	0.5895	0.5817	0.8794	0.8792	0.8794
Brightness, saturation, hue, contrast	0.5961	0.5995	0.5921	0.8918	0.8918	0.8918
Brightness, saturation	0.5786	0.5827	0.5738	0.8930	0.8929	0.8930

TABLE VI. RESULTS OF RESNET18 MODEL WITH DIFFERENT COLOR AUGMENTATION APPLIED FOR MULTICLASS CLASSIFICATION ON DATA WITH SHARED CLUSTERING WITH OMITTED INTERVALS BETWEEN CATEGORIES

Category split	First validation set			Second validation set		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Original	0.4968	0.3638	0.3585	0.9118	0.9111	0.9112
Brightness, saturation, hue, contrast	0.5464	0.3947	0.3892	0.9237	0.9216	0.9218
Brightness, saturation	0.5324	0.3823	0.3760	0.9213	0.9207	0.9213

[4] N. Naik, A. Madani, A. Esteva, N. S. Keskar, M. F. Press, D. Ruderman, D. B. Agus, and R. Socher, "Deep learning-enabled breast cancer hormonal receptor status determination from base-level H&E stains," Nat. Commun., vol. 11, no. 1, p. 5727, Nov. 2020.

[5] P. Seegerer, A. Binder, R. Saitenmacher, M. Bockmayr, M. Alber, P. Jurmeister, F. Klauschen, and K.-R. Müller, Interpretable Deep Neural Network to Predict Estrogen Receptor Status from Haematoxylin-Eosin Images. Cham: Springer International Publishing, 2020, pp. 16–37. [Online]. Available: https://doi.org/10.1007/978-3-030-50402-1_2

[6] R. R. Rawat, I. Ortega, P. Roy, F. Sha, D. Shibata, D. Ruderman, and D. B. Agus, "Deep learned tissue "fingerprints" classify breast cancers by er/pr/her2 status from h&e images," Scientific Reports, vol. 10, no. 1, p. 7275, Apr 2020. [Online]. Available: https://doi.org/10.1038/s41598-020-64156-4

[7] Y. Liu, X. Li, A. Zheng, X. Zhu, S. Liu, M. Hu, Q. Luo, H. Liao, M. Liu, Y. He, and Y. Chen, "Predict ki-67 positive cells in h&e-stained images using deep learning independently from ihc-stained images," Frontiers in Molecular Biosciences, vol. 7, 2020. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fmolb.2020.00183

[8] Y. Liu, T. Zhen, Y. Fu, Y. Wang, Y. He, A. Han, and H. Shi, "Ai-powered segmentation of invasive carcinoma regions in breast cancer immunohistochemical whole-slide images," Cancers, vol. 16, no. 1, 2024. [Online]. Available: https://www.mdpi.com/2072-6694/16/1/167

[9] F. Xing, H. Su, J. Neltner, and L. Yang, "Automatic ki-67 counting using robust cell detection and online dictionary learning," IEEE Transactions on Biomedical Engineering, vol. 61, no. 3, pp. 859–870, 2014.

[10] R. S. Geread, P. Morreale, R. D. Dony, E. Brouwer, G. A. Wood, D. Androutsos, and A. Khademi, "Ihc color histograms for unsupervised ki67 proliferation index calculation," Frontiers in Bioengineering and Biotechnology, vol. 7, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:202865235

[11] R. S. Gomolka, A. Korzynska, K. Siemion, K. Gabor-Siatkowska, and W. Klonowski, "Automatic method for assessment of proliferation index in digital images of dlbcl tissue section," Biocybernetics and Biomedical Engineering, vol. 39, no. 1, pp. 30–37, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0208521618302699

[12] T. Mungle, S. Tewary, I. Arun, B. Basak, S. Agarwal, R. Ahmed, S. Chatterjee, A. K. Maity, and C. Chakraborty, "Automated

[16] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, characterization and counting of ki-67 protein for breast cancer prognosis: A quantitative immunohistochemistry approach," Computer Methods and Programs in Biomedicine, vol. 139, pp. 149–161, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169260715302388

[13] D. Petríková, I. Cimrák, K. Tobiášová, and L. Plank, "Ki67 expression classification from he images with semi-automated computer-generated annotations," in BIOINFORMATICS, 2024.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Communications of the ACM, vol. 60, pp. 84 – 90, 2012.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," CoRR, vol. abs/1409.1556, 2015.
V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," 2015, pp. 1–9.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016.

[18] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255.

[19] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological image analysis: a review," IEEE Rev. Biomed. Eng., vol. 2, pp. 147–171, Oct. 2009.

[20] N. Madusanka, P. Jayalath, D. Fernando, L. Yasakethu, and B.-I. Lee, "Impact of H&E stain normalization on deep learning models in cancer image classification: Performance, complexity, and trade-offs," Cancers (Basel), vol. 15, no. 16, Aug. 2023.

[21] A. Madabhushi and G. Lee, "Image analysis and machine learning in digital pathology: Challenges and opportunities," Med. Image Anal., vol. 33, pp. 170–175, Oct. 2016.

[22] M. Runz, D. Rusche, S. Schmidt, M. R. Weihrauch, J. Hesser, and C.-A. Weis, "Normalization of HE-stained histological images using cycle consistent generative adversarial networks," Diagn. Pathol., vol. 16, no. 1, p. 71, Aug. 2021.

[23] D. Petríková, I. Cimrák, K. Tobiášová, and L. Plank, "Semi-automated workflow for computer-generated scoring of Ki67 positive cells from he stained slides." in BIOINFORMATICS, 2023, pp. 292–300.